

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2002, 23:36:27 ; Search time 5307.2 Seconds  
(without alignments)  
2006.538 Million cell updates/sec

Title: US-09-719-748-1\_COPY\_98\_886

Perfect score: 789  
Sequence: 1 tatgacacgagagagagct.....ctctcagacacccctgac 789

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estda: \*  
2: em\_esthm: \*  
3: em\_estln: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_estl: \*  
10: gb\_estc: \*  
11: gb\_hic: \*  
12: gb\_gss: \*  
13: em\_gss\_hum: \*  
14: em\_gss\_liv: \*  
15: em\_gss\_pln: \*  
16: em\_gss\_vrl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID       | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 562.8 | 71.3        | 1729   | 11 AK004350 | AK004350 Mus muscu |
| 2          | 541.4 | 68.6        | 882    | 9 A1322362  | A1322362 me98c10.y |
| 3          | 501.4 | 63.5        | 830    | 10 BG688701 | BG688701 602787522 |
| 4          | 494.6 | 62.7        | 976    | 11 BC012210 | BC012210 Mus muscu |
| 5          | 483   | 61.2        | 972    | 10 B1554559 | B1554559 603235693 |
| 6          | 464.2 | 58.8        | 1199   | 11 AK009701 | AK009701 Mus muscu |
| 7          | 459.2 | 58.2        | 869    | 10 W82116   | W82116 me98c10.r1  |
| 8          | 428.2 | 54.3        | 671    | 10 BF727181 | BF727181 By17h06.y |
| 9          | 425.6 | 53.9        | 982    | 10 BF976537 | BF976537 602244391 |
| 10         | 422.2 | 53.5        | 708    | 9 BG384579  | BG384579 303638 MA |
| 11         | 414   | 52.5        | 708    | 9 AM476323  | AM476323 uc73f11.y |
| 12         | 413   | 52.1        | 649    | 10 B1046467 | B1046467 M03-EN020 |
| 13         | 411.4 | 52.3        | 1127   | 10 BE733365 | BE733365 601567911 |
| 14         | 410.6 | 52.0        | 555    | 10 BF087508 | BF087508 QV2-HT054 |
| 15         | 409   | 51.8        | 1058   | 10 BG421064 | BG421064 602451186 |
| 16         | 408.2 | 51.7        | 663    | 10 BF019568 | BF019568 ux12f12.y |
| 17         | 403.4 | 51.1        | 565    | 10 BF075625 | BF075625 224820 MA |

|    |       |      |      |             |                    |
|----|-------|------|------|-------------|--------------------|
| 18 | 400.4 | 50.7 | 999  | 10 BF168866 | BF168866 601775325 |
| 19 | 394.4 | 50.0 | 556  | 10 BF075639 | BF075639 224840 MA |
| 20 | 394.4 | 50.0 | 696  | 10 BF012103 | BF012103 ux53901.y |
| 21 | 391.4 | 49.6 | 871  | 10 B1690902 | B1690902 603312045 |
| 22 | 386   | 48.9 | 681  | 10 B1149425 | B1149425 602848986 |
| 23 | 385.4 | 48.8 | 797  | 10 B1663166 | B1663166 603286787 |
| 24 | 385.2 | 48.8 | 887  | 10 BG968191 | BG968191 602835674 |
| 25 | 379   | 48.0 | 855  | 10 B1828782 | B1828782 603074924 |
| 26 | 378   | 47.9 | 703  | 10 BG65492  | BG65492 602783624  |
| 27 | 377   | 47.8 | 1039 | 10 BG419640 | BG419640 602451853 |
| 28 | 376.2 | 47.7 | 756  | 10 BF162798 | BF162798 601769262 |
| 29 | 372.4 | 47.2 | 460  | 9 AA858002  | AA858002 of63908.s |
| 30 | 371.4 | 47.1 | 654  | 10 B1012312 | B1012312 B0132312  |
| 31 | 370.4 | 46.9 | 796  | 10 BG421646 | BG421646 602449739 |
| 32 | 365.2 | 46.3 | 481  | 9 AM603538  | AM603538 RC0-CN002 |
| 33 | 363.8 | 46.1 | 710  | 10 BG277312 | BG277312 ux42b06.y |
| 34 | 358.8 | 45.5 | 785  | 10 B1652999 | B1652999 603299971 |
| 35 | 355.8 | 45.1 | 707  | 10 B1904738 | B1904738 603168651 |
| 41 | 341.6 | 43.3 | 589  | 9 AM421982  | AM421982 f189d06.y |
| 42 | 339   | 43.0 | 958  | 10 B1757256 | B1757256 603030761 |
| 43 | 338.8 | 42.9 | 743  | 10 BG968192 | BG968192 602835675 |
| 44 | 337.8 | 42.8 | 623  | 10 BE848999 | BE848999 ux01907.y |
| 45 | 326.8 | 41.4 | 1300 | 10 BF685010 | BF685010 602142870 |

#### ALIGNMENTS

|            |  |             |      |        |                 |
|------------|--|-------------|------|--------|-----------------|
| RESULT 1   | AK004350   | 1729 bp     | mRNA | linear | HTC 19-JAN-2002 |
| LOCUS      | AK004350   |             |      |        |                 |
| DEFINITION | Mus musculus 18 days embryo whole body cDNA, RIKEN full-length enriched library, clone:1110062102;death-associated kinase 2, full insert sequence.   |             |      |        |                 |
| ACCESSION  | AK004350   |             |      |        |                 |
| VERSION    | AK004350.1   | GI:12835501 |      |        |                 |
| KEYWORDS   | HTC; CAP trapper.  |             |      |        |                 |
| SOURCE     | Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA, clone:1110062102.  |             |      |        |                 |
| ORGANISM   | Mus musculus   |             |      |        |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  |             |      |        |                 |
| REFERENCE  | 1 (sites)  |             |      |        |                 |
| REFERENCE  | Carninci, P. and Hayashizaki, Y.   |             |      |        |                 |
| REFERENCE  | High-efficiency full-length cDNA cloning   |             |      |        |                 |
| REFERENCE  | Meth. Enzymol. 303, 19-44 (1999)   |             |      |        |                 |
| REFERENCE  | 99279253   |             |      |        |                 |
| REFERENCE  | 10349636   |             |      |        |                 |
| REFERENCE  | 2 (sites)  |             |      |        |                 |
| REFERENCE  | Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  |             |      |        |                 |
| REFERENCE  | Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes   |             |      |        |                 |
| REFERENCE  | Genome Res. 10 (10), 1617-1630 (2000)  |             |      |        |                 |
| REFERENCE  | 20499374   |             |      |        |                 |
| REFERENCE  | 11042159   |             |      |        |                 |
| REFERENCE  | 3 (sites)  |             |      |        |                 |
| REFERENCE  | Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, A., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. |             |      |        |                 |
| REFERENCE  | RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer  |             |      |        |                 |
| REFERENCE  | Genome Res. 10 (11), 1757-1771 (2000)  |             |      |        |                 |





QY 181 caccacatgcatcagcctcagcagcgtctatgaaacccgacgctgctgacatc 240  
 Db 359 CACCCCAACATCATCACCTCCACAGACGCTATGAGAACCCAGACGCTGCTCTATC 418  
 QY 241 ctgagcagctgctcctgagagcctctctgattctctgcccagaagagctgagt 300  
 Db 419 CTTGAGCTAGCTCCGAGGAGAACTGTTGATTTCTGCCCCAGAGAGTCTGTAAT 478  
 QY 301 gaggagagagccacacagctcattatgaagacagatcctgagtggttgacactcctcacaca 360  
 Db 479 GAGAGAGAGAGCCACACCTCATTAAGCAGATCCTGATGGGTGAATTAACCTTCACACA 538  
 QY 361 aagaataatgctcacttgatctcctcagcagaatacattatgtgta-gacaagaatat 419  
 Db 539 AAGAAATGCTCATTGATCTCAAGCCAGAAACATCATGTTGTTAGACACAAGATAT 598  
 QY 420 tccaccatcacacatcagctgattgacttgctgctgcccagaagaatgagatgga-g 478  
 Db 599 CCCCATTCACACATCAAGCTGATGACTTTGGCTGGCTGCACGAAATAGAAAGATGAGC 658  
 QY 479 tttaattgaagaattt-ttggagacgcggaattgtgtgtctccagaattgtgaactac 537  
 Db 659 TTGAATTTAAACATGTTTGGACACCTGATTTGTTGCTCCAGAAATCGTGAACATAT 718  
 QY 538 gaggccctgag-gctgagagctgacatgtgagacatagcgctcatcactacactcctt 596  
 Db 719 GAGCCACTGGGAGACTGGAGTGGACATGTGGACATTTGAGTCAATCACTTAATCTCTCT 778  
 QY 597 aagtgagacatccctcttccctggagagacagagacaggaacacatgagcaat 648  
 Db 779 AAGTGGAGGCTGCCCTTCTGAGAGACACAAACAGAAACCTTGCAATAT 829

RESULT 4  
 LOCUS BC012210 976 bp mRNA linear HTC 08-AUG-2001  
 DEFINITION Mus musculus, Similar to Death-associated like kinase, clone  
 ACCESSION BC012210  
 VERSION BC012210.1 GI:15126556  
 KEYWORDS HTC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Strausberg, R.  
 Direct Submission  
 Submitted (06-AUG-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: http://mgc.nci.nih.gov  
 Contact: MGC help desk  
 Email: gcgaps-remail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
 Contact: villalonedcm.tmc.edu.  
 Villalon, D.K., Luna, R.A., Hale, S.M., Huljk, S., Lu, X., Garcia,  
 A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
 Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov  
 Series: IRAP Plate: 23 Row: n Column: 16  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Genomescan gene

FEATURES  
 source  
 1..976  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3993055"  
 /tissue\_type="Mammary tumor. MAP-TGF alpha model. 7 months  
 old, gross tissue."  
 /clone\_id="NCL CGAP\_Mam5"  
 /lab\_host="DHI0B"  
 /note="Vector: pCMV-SPORT6"  
 BASE COUNT 240 a 271 c 294 g 171 t  
 ORIGIN

Query Match 62.7% Score 494.6 DB 11 Length 976;  
 Best Local Similarity 76.7% Pred No. 6.4e-109;  
 Matches 605; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 1 tatgacatcgagagagagctgaggagtgccagtttgcacatgtaagaagtcgagag 60  
 Db 131 TATGAGATGGAGAGAGAGACTTGCCAGCAATTTGGCATCGGCCAATGCGACAG 190  
 QY 61 aagaacaggggcttgagtagcagccaagtctatcaagaagcgagcgagcgagc 120  
 Db 191 AAGGCGAGGCGATGAGATATGACGCCAAGATTCTCAAGAAAGCGGCCCTGCGATCCAGC 250  
 QY 121 cggcgagtgtagcggagagagatcgagcgaggtgagatcctcgcgagcgagtgctg 180  
 Db 251 CGGCGGCTGTGAGACCCGGAGAGATCGAAGCGAGGTGAGCATCTCGCCGAGATCCG 310  
 QY 181 caccacatgcatcagcctcagcagcgtctatgagaaacccgacgctgctgacatc 240  
 Db 311 CACCCCAACATCATTAACATGACATGATGTTGAGAAACAGACAGATGTGTGATGATC 370  
 QY 241 ctgagcagtgctcctgagagagagctctcattctcctgcccagaagagagtagagct 300  
 Db 371 CTGAGCTGTGTGTCGGGTGCGAGACTTTTGACATCTCTGCGCCAGAAAGAGATATTGAGC 430  
 QY 301 gaggagagagccacacagctcattatgaagcagatcctgagtgaggtagaactcctcacaca 360  
 Db 431 GAGGATGAGGCGCACGCGATGCTCAAAACAAATCTGAGCGGTGTCACATCGACCTCC 490  
 QY 361 aagaataatgctcacttgatcctcaagcagaagaataatgattgttgaagaagaattt 420  
 Db 491 AAGCGATCGCACATCTTGACCTGGAAGCCCGAGAACATCATGTTGCTGSAACAACGCA 550  
 QY 421 cccattccacacatcaagctgattgacttgctgctcagcaagaatagaatgagtgat 480  
 Db 551 GCCAGCCCGCGCATTAAGCTCATCGACTTTGGCATCGCGCACAGAGTGAAGCTGGCAGC 610  
 QY 481 gaattaaagaatatttttggagcgcggaattgtgtgtccagaataatgtgaactagag 540  
 Db 611 GAGTTCAAAACATCTTTTGCGACACCCGAGTTGTGCGCCCGAGATGCTGAATATGAG 670  
 QY 541 cccctgggtctgagagctgacatgtgagacatagcgatcgaactaactaactccttaagt 600  
 Db 671 CCACCTGGCTTGAGAGCTGACATGTGAGAGCATTTGCGCTCATCACTTAATCTCTGAGC 730  
 QY 601 gggagatcccttctcctgaggagacagagaagaacactggaataatatacatcagtg 660  
 Db 731 GGAGGCTCCCATTCCTTGCGGAGACCAAGAGAGAGAGCGTGAAGCAATCTCGACAGTG 790  
 QY 661 agttagcacttgatgaggaattcttaagcattcagcgagcgagcgagcgaggaacttaatt 720  
 Db 791 AACTATGACTTTGATGAGAGATTAATTCTTCAAGCAACCGAGCGTGGCCAAAGACTTCATC 850  
 QY 721 cggaaagctcttgatgaagaagacccggaaagagctcacaacccaagaagcgctcgaagcac 780  
 Db 851 CGAGGCTGCTGTCAAAAGACCCCAAGAGAGAGATGACATCGACAGAGCTGAGACTAT 910  
 QY 781 cccctgagatc 789

Db 911 TCCGCGATC 919

RESULT 5  
BI554559 972 bp mRNA linear EST 05-SEP-2001  
LOCUS 603235693F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:5310046 5',  
DEFINITION mRNA sequence.  
ACCESSION BI554559.1 GI:15441873  
VERSION BI554559.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 972)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: L19M1785 row: 1 column: 23  
High quality sequence stop: 845.  
Location/Qualifiers  
Source 1..972  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5310046"  
/clone\_1lb="NCI\_CGAP\_L19"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: Liver; Vector: PCMV-SpRt6; Site\_1: NotI;  
Site\_2: SalI; cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 239 a 269 c 298 g 166 t  
ORIGIN

Query Match 61.2%; Score 483; DB 10; Length 972;  
Best Local Similarity 77.5%; Pred. No. 4e-106;  
Matches 585; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 1 tatgacatcggagagagcgtgaggagtcgacattcgccatcgtaagaagtcgcggag 60  
|||||  
Db 32 TATGACATGAGAGAGAGAGCTTGCGACATTTGCCATTCGTGCCAAGTCCACAG 91  
|||||

QY 61 aagagcagcgagcgttgatgacgcaagtcctcaagaagcgagcgagcgagcgagc 120  
|||||  
Db 92 AAGGCGACGGGATGAGATGACGCCAAGTTCATCAAGAAAGCGCGCTCCATCCAGC 151  
|||||

QY 121 cggcgagtcgtgagcggagagagatcgagcgagagtgacatcctcgcgagtcgtg 180  
|||||  
Db 152 CGGCGCGGTGTGACCCGCGAGAGATGAGACGAGGTGACATCTCTCGCGAGATCCG 211  
|||||

QY 181 caccacaatgcatcagcgtgacgacgctctatgagaacgcgacgcgagtcgtgacatc 240  
|||||  
Db 212 CACCCCAACATCATACATGACATGACGTTCGACGAACAAGACAGATGTGTCTGATC 271  
|||||

QY 241 cttagagctagtcgtgagagagcctctgattcctcgcccgagagagagtcagtcagt 300  
|||||  
Db 272 CTGAGAGCTGGTGTCCGGGCGGAGCTTTTGCACCTTCCTGCGCGAAGAGATCATTTGACG 331  
|||||

QY 301 gaggagagagcgacacagctcatlaagcagatcctgagatggggtgaactacatcaacaca 360  
|||||

Db 332 GAGGATGAGGGCCAGCAGTTCCTCTCAACAATCTAGACGGTGTTCACCTACCTGACTCC 391

QY 361 aagaaatgctcaacttgatctcaagccagaataatgattgtttagaagaatatt 420  
|||||

Db 392 AAGGCGATCGCACACTTGTGACCTGAGCCGAGAACATCATGTGTGGACAACACGCA 451  
|||||

QY 421 cccatccacatcaagctgattgacttgctgctgcgcgaataatgagagagatt 480  
|||||

Db 452 GCCAGCCCCCGCATTAACCTCATGCTATCTTGGCATCGCGCAGAGATGAGGCTGGCAGC 511  
|||||

QY 481 gaattagaatatttttgagacgcgggaatttctgtctccagaatattgtaactagag 540  
|||||

Db 512 GAGTTCAAGAACATCTTTGGCACACCCGAGTTTGTGGCCCCGAGATGTGTAACATAG 571  
|||||

QY 541 cccctgggtctggaagctgacatgtgagcaatagagcgtacatcactacatccttaagt 600  
|||||

Db 572 CCACCTGGCTTGGAGGCTGACATGTGGAGCATTTGGCGCTCATCTACCTCTCTGAGC 631  
|||||

QY 601 gggagatcccttctccttgaggagacagagagaggaacatcggaataatcacatcagt 660  
|||||

Db 632 GAGCGCTCCCATTTCTGGCGAGACCAAGCAGAGAGCTGACGACATCTCAGACAGTG 691  
|||||

QY 661 agttagcatttgatgaggaattcttcagccatacagagagctgtgcccagaacttatt 720  
|||||

Db 692 AACATGACTTTGATGAGGATTAATCTTACGACAGCAGCAGCTGCGCAAGACTTCATC 751  
|||||

QY 721 cggagctctcgtttaaagagacccggaacgct 755  
|||||

Db 752 CGCAGGCTGCTGTCAAGACCCCAAGAGAGAGAT 786  
|||||

RESULT 6  
AK009701 1199 bp mRNA linear HTC 19-JAN-2002  
LOCUS Mus musculus adult male tongue cDNA, RIKEN full-length enriched  
DEFINITION library, clone:2310039H24:DEATH ASSOCIATED PROTEIN KINASE, full  
insert sequence.  
ACCESSION AK009701 GI:12844657  
VERSION AK009701.1  
KEYWORDS HTC; GAP trapper.  
SOURCE Mus musculus (Strain:C57BL/6J) adult male tongue cDNA to mRNA,  
clone\_1lb:RIKEN full-length enriched mouse cDNA library  
clone:2310039H24.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (sites)  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (sites)  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 11042159  
PUBMED 20499374  
REFERENCE 3 (sites)  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, D., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multipillar sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="b17h06"
/clone_lib="Human Lens cDNA (un-normalized, unamplified):
By"
/tissue_type="Lens"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pcwmsport6; Two human lenses
from different adults (both approximately 40 years old)
together yielded 20ug of total RNA and 150ng mRNA for cDNA
library synthesis. A directionally cloned cDNA library in
the pcwmsport6 vector was constructed at Life Technologies
, essentially following the protocols of the Superscript
Plasmid System full details of which are contained in the
manufacturer's instruction manual
(http://www.lifetech.com/). First strand synthesis was
carried out using a Not I primer-adaptor
[5'-GACATGATTCCTACATCGACGCGCCGCTT)15-3']. Not I/blunt
end inserts were cloned into the Not I/EcoR V sites in the
vector. EST analysis was performed on the unamplified
library at the NIH Intramural Sequencing Center (NISC)."

```





```

Db 541 GAGACCCAGAGAGAGAGCTCACCACATCTCAGCCCTGACACTGACACTTCGAGAGAGAG 600
Oy 682 tcttcagcacaacagagagcgtgccaagacttactcggaagcttcgtttaagag 741
Db 601 TACTTCAGCAACACACAGAGCTGGCCAGAGACTTCATTCGCCGCTGCTCCGCAAGA 660
Oy 742 accc 745
Db 661 TCCC 664

RESULT 10
LOCUS BG384579 585 bp mRNA linear EST 12-MAR-2001
DEFINITION 303638 MARC 1Pig Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BG384579
VERSION BG384579.1 GI:13308988
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE 1 (bases 1 to 585)
AUTHORS Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.M.
and Keele, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -minscore 18
and -mismatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGATCAGCAGC
Plate: 91 row: M column: 15
Seq primer: ATTGAGTGCACCTATAG.
FEATURES
Source
1. 585
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_1lb="MARC 1Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; site_1: XbaI; site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 137 a 159 c 178 g 111 t
ORIGIN
Query Match 53.5%; Score 422.2; DB 10; Length 585;
Best Local Similarity 94.0%; Pred. No. 1.6e-91;
Matches 439; Conservative 0; Mismatches 28; Indels 0; Gaps 0;
Oy 1 tatgacatcgagagagcgtggagagtcgcaagcttcgcatcgtgaagaagtcgagag 60
Db 119 TACACATCTCGGAGAGAGCTGGGAGCGGCCAGTTCACATCGGAAGAGTCCCGGAG 178
Oy 61 aaagagcggagcgttgagtgagcgaagtcataagaagcggcagagcggcgagc 120
Db 179 AAGAGCAGGGAGCTGGAGTTCAGTTCATCAAGAGCGGAGCGGCGCAGC 238
Oy 121 cggcgagtgatgagcggagagagatcgagcggagagtgagacatccctgagcgagtgctg 180
Db 239 CGCGGGGGCGTGTCCCGCGGAAGATTTGAGCGGAGGTGGCATCTGCGCGAGTCTG 298

```

```

Oy 181 caccacaatgtcatcagcgtgacagagtgctatagaaacgcaccagcgtgtgacatc 240
Db 299 CACCCCAAGCTCATCAGCTGCGACGAGCTCTTTGAGAACCGCACCGAGCTGCTCATC 358
Oy 241 cttagctagtgctgagagagagccttcgattccctgagcccaagaagagacagtgag 300
Db 359 CTCGAGCTAGTGTCTGAGAGAGAGCTCTTCGACTTCTCGGCCCAAGAGACTCCCTGAGT 418
Oy 301 gagagagagcaccagcttcattgaagcagatcctgagatgggtgaactacattcacaca 360
Db 419 GAGAGAGAGCGGACAGCAGCTTCATTAGACAGATCCTGATGGGTGAGACTACCTTCATGCC 478
Oy 361 aagaattgtcctcatttgatcctcaagccagaataatagtggttagaagaataatc 420
Db 479 AAGAAATTTGCTCACTTTGACTCAAGCCAAAAACATTAAGTTGTTGGAGAAAGATATT 538
Oy 421 cccattcacacatcaagcgtgacttgctgtgctgagcagaaat 467
Db 539 CCCATTCCACATCAAGCTGATTGATTTGGCTGCTGCTCATGAAT 585

RESULT 11
LOCUS AM476323 708 bp mRNA linear EST 24-FEB-2000
DEFINITION uq73f11.y1 NCI-CGAP_Lu33 Mus musculus cDNA clone IMAGE:2936781 5'
sequence.
ACCESSION AM476323
VERSION AM476323.1 GI:7046429
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 708)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other-ESTs: uq73f11.x1
Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/dbfp/image/image.html
MGI:1049201
Seq primer: -40RP from G1bco
High quality sequence stop: 471.
FEATURES
Source
1. 708
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:2936781"
/clone_1lb="NCI-CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCCTGTTGTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was

```

BASE COUNT 169 a 205 c 205 g 128 t 1 others  
 ORIGIN constructed by Bento Soares and M. Fatima Ronaldo.

Query Match 52.5% Score 414; DB 9; Length 708;  
 Best Local Similarity 75.1%; Pred. No. 1.6e-89;  
 Matches 516; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 103 cggagagccggcgccggtgtgagccggagagagatcgaaggagtgagc 162  
 Db 1 CGGGCGCTGCTTCCAGCCGGCGGTGTGAGCCGGAGAGATCAGACCGCGAGAGC 60  
 QY 163 atctcgagcaggctgtgacacacatgtatcagcgtgacagcgtctatagagaccg 222  
 Db 61 ATCTCGGCGAGATCCGCCACCCCAACATCATMACTGATGACGTGTTCCGAAACAAG 120  
 QY 223 accgagctgtgacacatcctgtgagctgtgagagagagctcttcgattcctgccc 282  
 Db 121 ACAGATGTGCTGTGATCTCTGGAGCTGTGTCGGTGGGAGCTTTTCGACTTCTGGCC 180  
 QY 283 cagaagagtgactgagtgagagagagagcaccagctcattatagagagctgtgag 342  
 Db 181 GAGAAAGAGATGATGACGAGAGATGAGGACGACATTCCTCAACAAATCTAGACGT 240  
 QY 343 gtgaactacttcacacaaagaattgtcacttgcattcgaagccagaagaattag 402  
 Db 241 GTCCACTACTGCACTCCCAAGCGCATCGCACATTTGACCTGAAGCCGCAACATCATNG 300  
 QY 403 ttgttagacaagaatattccattccacacatcagcgtgattgtgtgtgtac 462  
 Db 301 TTGCTGACAGAGCAGCAGCAGCCGCCCATTAAGCTATGACTTGTGCAATCGCAC 360  
 QY 463 gaaatagaagatgagatgtaattagaatattttgagagccggaatttgttgcac 522  
 Db 361 AGAGTCGAGGCTGGCAGCGCAGATTCAGAAACATTTGGCACCCGAGTTGTGCCCCC 420  
 QY 523 gaaatgtgaactacagagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 582  
 Db 421 GAGATGCTGATATGAGCACTTGTGAGGCTGACATGTGAGATGAGATGAGCTGTATC 480  
 QY 583 acctacatccttgaatgagagatccttctgtgtgtgtgtgtgtgtgtgtgtgtgt 642  
 Db 481 ACCTCATCTCTCTGAGGCGAGCGCTCCCATTTCTGTGGCGAGACACAGAGACGCTG 540  
 QY 643 gaaatatacatcagtgatgagtgagtgagtgagtgagtgagtgagtgagtgagtg 702  
 Db 541 ACGAACATCTGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 QY 703 ctgagcgaagacttatctggaagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 762  
 Db 601 CTGGCCAAAGACTTTCATCGCAGGCTGTGTCAAAAGACCCCAAGAGAGATGACCATC 660  
 QY 763 caagaggtctcagaccacctgtatc 789  
 Db 661 GCACAGAGCTTGAACATTCCTGGATC 687

RESULT 12  
 B1046467 649 bp mRNA linear EST 14-JUN-2001  
 LOCUS B1046467/c  
 DEFINITION M3-FN0209-070201-010-f01 FN0209 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION B1046467  
 VERSION B1046467.1 GI:14453089  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 649)  
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 2020263  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&t2=MR3-FN0209-  
 070201-010-f01&t3=2001-02-07&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 22  
 High quality sequence stop: 648.  
 Location/Qualifiers  
 1. 649  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="FN0209"  
 /dev\_stage="Adult"  
 /note="Organ: prostate-normal; Vector: puc18; Site\_1: SmaI  
 ; Site\_2: SmaI; A mini-library was made by cloning  
 products derived from ORESTES PCR (U.S. Letters Patent  
 Application No. 196,716 - Ludwig Institute for Cancer  
 Research) profiles into the puc 18 vector. Reverse  
 transcription of tissue mRNA and cDNA amplification were  
 performed under low stringency conditions."

FEATURES  
 source  
 Location/Qualifiers  
 1. 649  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="FN0209"  
 /dev\_stage="Adult"  
 /note="Organ: prostate-normal; Vector: puc18; Site\_1: SmaI  
 ; Site\_2: SmaI; A mini-library was made by cloning  
 products derived from ORESTES PCR (U.S. Letters Patent  
 Application No. 196,716 - Ludwig Institute for Cancer  
 Research) profiles into the puc 18 vector. Reverse  
 transcription of tissue mRNA and cDNA amplification were  
 performed under low stringency conditions."

BASE COUNT 151 a 162 c 160 g 176 t  
 ORIGIN

Query Match 52.3% Score 413; DB 10; Length 649;  
 Best Local Similarity 94.5%; Pred. No. 2.7e-89;  
 Matches 483; Conservative 0; Mismatches 20; Indels 8; Gaps 5;

QY 282 ccagaagagtcactggtgtagagagagccacagctcattcaagaagatcctgtatg 341  
 Db 636 CCCAGAAGGATCACTAGTAGAGAGGA-GCCACACAGTTTATTAACAGATCCCTGATAG 578  
 QY 342 gttgaactacttcacacaaagaattgtcacttgcattcgaagccagaagaattat 401  
 Db 577 GGTGAATCTTACACCAAAAGAAATGCTCACTTGTGATCTCA---AGAAAACATTTAT 522  
 QY 402 gtgttagacaagaatattccattccacacatcagcgtgattgtgtgtgtgtgtgtgt 461  
 Db 521 GTTGTATACAGATATATTCAGGCGACACATCAAGCTGATTTGCTGTGCTCA 462  
 QY 462 cgaatagaagaatgagtgtaattagaatattttggaagcgggaatttgttctcc 521  
 Db 461 CGAAATGAGAGATGAGTGAATTAATTTTGGAGCCCGGAATTTGTTCTCTCC 402  
 QY 522 a-gaaatttgaactacagagccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 580  
 Db 401 ACGAATTTGTGAATCGAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 342  
 QY 581 tcaactacatctc- ttaagttagagatcccttctcgtgtgtgtgtgtgtgtgtgtgt 638  
 Db 341 TCACCTACATCTCTTTAAGTAGAGATCCCTTCTCTGTGTGTGTGTGTGTGTGTGT 282  
 QY 639 acttggaatatacatcagtgatgagtgagtgagtgagtgagtgagtgagtgagtgag 698  
 Db 281 ACTGGCAATATCAAGCAGTAGTATGACATTTGATGGAATTTCTACGACCAAGAG 222  
 QY 699 cgagctgccaagagcttatctggaagcttctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 758

Db 221 CGAGTCGCGCCAGAGACTTTATTCGGAAGCTTCTGTTAAAGAGACCCGGAAAGGCTCAC 162

QY 759 aatccaaaggagctctcagacacccctggatc 789

Db 161 AATCCAAAGAGCTCTCAGACACCCCTGGATC 131

RESULT 13

LOCUS BE733365 1127 bp mRNA linear EST 15-SEP-2000

DEFINITION 601367911F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3842667 5', mRNA sequence.

ACCESSION BE733365

VERSION BE733365

KEYWORDS BE733365.1 GI:10147357

SOURCE EST.

ORGANISM human.

REFERENCE Homo sapiens

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE NIH-MGC http://mgi.ncl.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@rs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: Image.Lnl.gov  
Plate: L1CM536 row: e column: 04  
High quality sequence stop: 723.

FEATURES

source 1. .1127  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3842667"  
/clone\_1lb="NIH\_MGC\_21"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Placenta; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 268 a 305 c 357 g 197 t

ORIGIN

Query Match 52.1%; Score 411.4; DB 10; Length 1127;  
Best Local Similarity 74.1%; Pred. No. 8, 1e-89;  
Matches 534; Conservative 0; Mismatches 186; Indels 1; Gaps 1;

QY 1 tatgacatcggagagagagctgaggagtgccagcttcgacatcgtaagaagtcgagggag 60

Db 184 TATGACATGCGGGAGAGCTGGGACGGCGGACAGTTTCGATCGTGGGAAAGTGGCGGAG 243

QY 61 aagaagcaggggcttgagtgatcagcagcaaatcatcaagaagggagagagcggggagc 120

Db 244 AAGGGCAGGGGAGGAGTACGACGACCAAGTTCATCAAGAGCCGCCCTGTCTATCCAG 303

QY 121 cggcgagtgagcagcgagagagatcgagcgagagtgagcatcctcgcgagtgctg 180

Db 304 CGGCGTGGGTGAGCCGCGAGAGATCGAGCGGAGGTGAACATCTCGCGGAGATCCGG 363

QY 181 caccacaatgtcaccagctgcagcagcgttatatgagaacccgacccgagcgtggtcacatc 240

Db 364 CACCCCAACATCATACACCTGACAGCATCTTCGAGAAACAAGAGCGAGCTGTCTCATC 423

QY 241 ctgagctagtgctcaggagagagctctcagcttccttcctggccccaagaagagctcag 300

Db 424 CTGAGAGTGGTCTCTGGCGGGAGACTCTTTCAGCTCTCTGGCGGAGAGAGTGGCTGACG 483

QY 301 gaagagagagccaccagctcattcaagacagatccctgtagtgggtgtagacttcacaca 360

Db 484 GAGAGCAGAGCCACCCAGTCTCTCAACAGATCCTCGAGCGGCGTTCACCTCAGCTCT 543

QY 361 aagaagaatgctcacttgatctcaagccagaagaactatgtgttagaagaagaatc 420

Db 544 AAGGCATCGCAGCTTTGAGACCTGAAGCCGAAAAACATCATGCTGGACAAAGACGTG 603

QY 421 cccatccacaacatcaagcagatgacttgctgctcagcagaaatagaagaatgag 480

Db 604 CCCAACCCAGCATCAAGCTCATGAGTCTCGGATCCCGCAGACAGATCGAGCGGGAGAC 663

QY 481 gaattagaataatttggagcgcggaatttgctccagaatltgtaactacagag 540

Db 664 GAGTTCAAGAACATCTTCGCGACCCCGAGTTTGTGGCCCGAGAGATTGTAACATAGAG 723

QY 541 cccctggcttgagagctgagcagatgtgagcagatagagcgtcatcactcactcctta 600

Db 724 CCGCTGGGCTCGAGCGGAGCATGTGAGCATCGTGTATCATCTTATCTCTGAGGG 783

QY 601 ggaagcatcccttcctcgtggagagacagagaggaagaacactgggaatatcatcag 660

Db 784 TGAATTCCTCGGTACCTTGGAGACCAAGGACAGAGAGCTCACAGATTCAGCGCT 842

QY 661 agttacagacttgatgaggaattctcagccatcagcagagctggcgaaggaattatc 720

Db 843 AACAGACACTTCGACAGAGATCTTAGAAGAACCAAGCTAAGTTGGCGAGAGGTTATC 902

QY 721 c 721

Db 903 c 903

RESULT 14

LOCUS BF087508/c 555 bp mRNA linear EST 19-OCT-2000

DEFINITION QV2-HT0541-130900-364-d03 HT0541 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF087508

VERSION BF087508.1 GI:10893218

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 555)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=6tz-QV2-HT0541-130900-364-d03&tz=2000-09-13&tz=1)  
Seq primer: puc 18 forward  
High quality sequence start: 4

High quality sequence stop: 555.

Location/Qualifiers

1. 555

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HM0541"

/dev\_stage="Adult"

/note="Organ: head,neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 127 a 140 c 134 g 154 t

ORIGIN

Query Match 52.0%; Score 410.6; DB 10; Length 555;

Best Local Similarity 99.0%; Pred. No. 9.7e-89;

Matches 413; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 373 cacttgcattcacaagcaaaacattatgtttagacaagaatatcccatccacac 432

DB 555 CACTTGATCTCAAGCCAGAAACATTATGTTAGACAAGATATCCCATCCACAC 496

QY 433 atcaagctgattgacttgcctgctgcacgaataagaatgagttgaattaaagat 492

DB 495 ATCAAGCTATTTGACTTGTGCTGCTCAGAAATAGAAAGATGAGTTGAATTAAGAAAT 436

QY 493 attttgagagcgcgaattgttgcctcagaataatgtgaacagagccctgggtc 552

DB 435 ATTTTGGGACGGCGAATTTGTTGCTCCAGAAATTTGTACACGACGCCCTGGTCCG 376

QY 553 gaggctgacatgttgagcatagagcgctacacacacacacacacacacacac 612

DB 375 GAGGCTGACATGTGAGCATAGCGCTCATACCTACATCCTTTAGTGGAGATCCCT 316

QY 613 ttccctgagagacagaagaagaacacatgcagcaaatatcacatcagttgaagact 672

DB 315 TTCCTGGGAGACACAGAGAGAAACACAGCAAAATATACAGCACTGATGACTTT 256

QY 673 gatgagaaattcttcagccatacagagcgagctgcccagaagacttattcggaagct 732

DB 255 GATGAGGAATTTCTTCAGCCAGACGAGCGAGCTGGCAAGACTTTATTGGAAAGCTTC 196

QY 733 gftaaagagaccggaaagagctcaatcaagaagagctctagaacacccctgagat 789

DB 195 GTTAAAGAGACCCGGAACGGCTCACATCCAAAGAGGCTCTCAGACACCCCTGATC 139

RESULT 15

LOCUS BG421064 1058 bp mRNA linear EST 14-MAR-2001

DEFINITION 602451186f1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4589709 5',

ACCESSION BG421064

VERSION BG421064.1 GI:13327479

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

JOURNAL NIH-MGC http://mgc.ncl.nih.gov/.

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaaps-remail.nih.gov

Tissue Procurement: DCM/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LLM323 row: k column: 22

High quality sequence stop: 615.

Location/Qualifiers

1. 1058

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4589709"

/clone\_lib="NIH-MGC\_14"

/tissue\_type="renal cell adenocarcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: kidney; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 295 a 278 c 339 g 146 t

ORIGIN

Query Match 51.8%; Score 409; DB 10; Length 1058;

Best Local Similarity 77.6%; Pred. No. 3e-88;

Matches 507; Conservative 0; Mismatches 145; Indels 1; Gaps 1;

QY 40 atcgtgaagaagtgcgcggagaaagacagcggtgtgagctgacgaacgaatcataag 99

DB 2 ATCGTCGGAGAGCGCGGAGAGGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 61

QY 100 aagcggcagagacggcgcgcggtgtgagcgggagagagatcgagcgagaggtg 159

DB 62 AAGCGCGCGCTGTCATCCAGCGCGCGGTGGGTGAGCGCGGAGAGATCGAGCGGAGGTG 121

QY 160 agcatctcgcgcgaggtgtgtgcacacaaatgtacagctgtgacagagctctatgaac 219

DB 122 AACATCTCGCGGAGATCCGGGACCCCAACATCATACCTGACACATCTTTCGGAAC 181

QY 220 cgcacccgagctgtgtgacacatccttgagctagtgctgtgagagagagctcttcga 279

DB 182 AAGAGGAGAGGTGCTCTCTCTGAGCTGAGCTGTGGGGGAGGCTCTTTCGACTTCG 241

QY 280 gcccaagagagctgacgtgagtgagagagagcaccacagcttcaatgaagatcccgat 339

DB 242 GCGGAGAGAGAGTCCGTGACGAGAGAGGCGCCACCCACTTCTCAAGCAGATCCGTGAC 301

QY 340 ggggtgaactacatccacacaaagaatgtctacatttgatctcaagccagaanaacatt 399

DB 302 GCGCTTCACTACCTCCTCACTTAAGCGCATCGCACACTTTCGACTGAAAGCGGAAACATC 361

QY 400 atgtgttagaagaagaatattccatctcacacatcaagctgattgacttgcgtgct 459

DB 362 ATGCTGCTGAGAACAGAGGTGCCAACCCCAATCAAGCTGATGACTTCGGCATTCGG 421

QY 460 cagcaaatagaagaatgagattgaattaaatatatttgaggagccgggaatttttgt 519

DB 422 CACAGAGATCGAGGCGGAGACAGATTCAAGAACTTCGCGACCCCGAGTTAGTGGCC 481

QY 520 ccagaatattgaaactacagacccctgggtctgagagctgacatgtgagacataagcgctc 579

DB 482 CCAGAGATTGTGACTATGAGCCCGCTGGGCTGAGGCGGAGCATGTGGAGCATGGTGTG 541

QY 580 atcaactacatccctttagagtgagagacatcccttccctgggagagacagaagaagaaca 639

DB 542 ATCACTATATCTCTCTGAGGCGGTGATCCCGTTCCTGCGGAGATC-AAGCAGAGAGCG 600

QY 640 ctggcaaatatcacatcagtgagttgagacttgatgagagatctcttaagca 692

DB 601 CTCACCAAGATCTCAGCGGTGAACTACGAAAGTCAGAGAGAGTACTTACGAA 653

Sat May 18 15:09:07 2002

us-09-719-748-1\_copy\_98\_886.rst

Page 13

Search completed: May 17, 2002, 23:36:36  
Job time: 5495 sec

---

This Page Blank (uspto)